DeepAge: Harnessing Deep Neural Network for Epigenetic Age Estimation From DNA Methylation Data of Human Blood Samples

Sajib Acharjee Dip[1], **Da Ma**[2], **Liqing Zhang**[1*]
[1]Department of Computer Science, Virginia Tech
[2]Wake Forest University School of Medicine
sajibacharjeedip@vt.edu, dma@wakehealth.edu, lqzhang@vt.edu

Abstract

Accurate prediction of biological age from DNA methylation data is a critical endeavor in understanding the molecular mechanisms of aging and developing age-related disease interventions. Traditional epigenetic clocks rely on linear regression or basic machine learning models, which often fail to capture the complex, non-linear interactions within methylation data. This study introduces DeepAge, a novel deep learning framework utilizing Temporal Convolutional Networks (TCNs) to enhance the prediction of biological age from DNA methylation profiles using selected CpGs by a Dual-Correlation based apparoach. DeepAge leverages a sequence-based approach with dilated convolutions to effectively capture long-range dependencies between CpG sites, addressing the limitations of prior models by incorporating advanced network architectures including residual connections and dropout regularization. The dual correlation feature selection enhances our model's predictive capabilities by identifying the most age-relevant CpG sites. Our model outperforms existing epigenetic clocks across multiple datasets, offering significant improvements in accuracy and providing deeper insights into the epigenetic determinants of aging. The proposed method not only sets a new standard in age estimation but also highlights the potential of deep learning in biologically relevant feature extraction and interpretation, contributing to the broader field of computational biology and precision medicine.

Introduction

In the realm of biomedical research, the accurate estimation of biological age from epigenetic data, specifically DNA methylation, represents a pivotal challenge and opportunity. Biological age, as opposed to chronological age, offers a more nuanced view of an individual's health status and aging process, informed by the epigenetic modifications that accumulate over time. These modifications, particularly methylation of DNA at CpG sites, have been robustly associated with various age-related changes and conditions. Traditional methods for estimating epigenetic age leverage linear and basic machine learning models, which, while foundational, often struggle with the complex, non-linear relationships inherent in methylation data across diverse biological systems. Various models such as Hannum (Hannum et al. 2013), Horvath1 (Horvath 2013), Horvath2 (Horvath et al. 2018), Lin (Lin et al. 2016), PhenoAge (Levine et al. 2018), and others employ diverse computational strategies, predominantly linear regression-based approaches that focus on weighted sums of methylation levels at selected CpG sites. These traditional models, while foundational, often do not account for the complex, non-linear interactions between CpGs that might influence aging processes more profoundly. For instance, Horvath's clocks use linear algorithms that may not capture the entire spectrum of biological aging changes, leading to limitations in prediction accuracy across diverse populations and tissues. There are some non-linear parametric regression based methods for example, GP-Age (Varshavsky et al. 2023) which improve over flexible prediction but still suffer from capturing complex interaction.

Recent advancements have seen the application of more sophisticated machine learning techniques, such as random forests (Breiman 2001) and gradient boosting machines (Chen and Guestrin 2016), which have provided incremental improvements and more flexibility over linear models. However, these methods still often fall short of capturing the deeper interactions within methylation profiles. There are some methods applying deep learning appoaches, for example, PerSEClock (Zhao et al. 2024) applied channel attention, CPFNN (Li et al. 2021) used correlation pre-filtered neural network, MSCAP (Wang et al. 2023) used multi-scale CNN model, but most of them lack in considering their sequential or collective influence on aging that can be critical for a precise age prediction. Furthermore, most existing models have not fully explored the potential of deep learning techniques, which have revolutionized fields such as image and speech recognition for detecting intricate patterns in high-dimensional data.

Our work introduces DeepAge, a novel deep learning framework specifically designed to address these challenges in the context of epigenetic age estimation. DeepAge utilizes Temporal Convolutional Networks (TCNs) (Lea et al. 2016), which are particularly adept at handling sequence data, to model the sequential nature of CpG sites across the genome. This approach allows for an effective capture of long-range dependencies and interactions between CpG sites, which are essential for understanding the complex biological processes underlying aging. By integrating layers

of temporal blocks that include dilated convolutions (Yu and Koltun 2015), DeepAge can access a broader context of the input sequence, thus enhancing its ability to discern pertinent aging signals from the methylation patterns. We implemented a dual correlation technique, utilizing both Spearman and Pearson correlations to identify CpGs most associated with age (De Winter, Gosling, and Potter 2016). By focusing on these relevant features, our model's performance improved significantly, reducing the risk of overfitting and enhancing generalizability.

Moreover, DeepAge incorporates advanced techniques such as residual connections (He et al. 2016) and dropout regularization (Srivastava et al. 2014) to refine its learning process and avoid overfitting, a common challenge in deep learning models dealing with high-dimensional biological data. The architecture is designed to progressively increase its receptive field without inflating the model size excessively, making it both powerful and computationally efficient. This allows DeepAge not only to outperform existing epigenetic clocks (Hannum et al. 2013; Horvath 2013; Horvath et al. 2018; Lin et al. 2016; Belsky et al. 2022) in terms of prediction accuracy but also to provide deeper insights into the epigenetic factors that drive biological aging.

In summary, DeepAge represents a significant step forward in the field of epigenetics, offering a robust, scalable, and interpretable tool for age estimation that leverages the full potential of deep learning. Our extensive evaluations across diverse datasets demonstrate its superior performance and underscore its potential to enhance our understanding of aging and its biological underpinnings, paving the way for improved diagnostic and therapeutic strategies in age-related diseases.

Materials and Methods

Dataset

For this study, we utilized 12 publicly available GEO datasets (Edgar, Domrachev, and Lash 2002) from the Biolearn library (Ying et al. 2023), encompassing a comprehensive age range from newborns to centenarians. The datasets provide rich metadata, including age and sex, allowing for a nuanced analysis of methylation patterns across different demographics. We partitioned these datasets into three groups: 90% of those were kept for training and validation while the remaining samples served as held-out test datasets to evaluate generalizability.

The age distribution across these datasets is depicted in the accompanying Fig. 1a, highlighting the mean and standard deviation of ages, which span from 0 to over 100 years. Notably, of the 12 datasets, 8 contain gender information, with a demographic composition of 69% male and 31% female samples, as illustrated in Fig. 1b. This gender representation ensures that our findings are robust across both male and female cohorts. Most samples originate from human whole blood tissue, which is commonly used in epigenetic studies due to its accessibility and the abundance of methylation data it offers. This tissue type enhances the relevance and applicability of our study to general and clinical research.

Data Preprocessing

The preprocessing of our dataset was uniformly applied across the training, validation, and testing sets to ensure consistency. For the independent test set, models were evaluated directly without additional preprocessing to assess their performance on unaltered data. We initially removed samples with over 50% missing methylation values and excluded any lacking precise age information or containing NaN values. To mitigate the impact of age outliers, samples older than 100 years were removed due to their insufficient numbers and potential to skew the model's learning.

The resulting dataset comprised 4,351 samples, each characterized by 20,937 CpG sites. We then normalized the methylation beta values to a 0 - 1 range using the MinMaxScaler from the scikit-learn library (Pedregosa et al. 2011), ensuring that all values were appropriately scaled for effective model input. Lastly, missing values within the methylation data were imputed using the mean methylation level of each CpG site across all samples, facilitating a consistent dataset for subsequent analyses. This rigorous preprocessing pipeline was essential for preparing the data for accurate and unbiased epigenetic age prediction.

Age Associated Feature Selection and Data Partitioning

In our study, we employed a novel Dual-Correlation feature selection technique that improved our model's predictive accuracy by integrating both Spearman's rank and Pearson's correlations to detect monotonic and linear relationships, respectively. By identifying significant CpG sites from both methods, we captured a comprehensive set of biomarkers indicative of epigenetic aging. This approach not only enhanced the robustness of our feature selection but also deepened our understanding of epigenetic changes with age. For correlation thresholds set at T=0.4,0.45,0.5,0.55, and 0.60, we identified 407, 184, 57, 16, and 5 CpG sites, respectively (Selection process is shown in the Fig. 3a for T=0.45). These varying thresholds allowed us to assess the impact of feature granularity on model performance, ultimately aiding in the optimal selection of predictive biomarkers.

Regarding data partitioning, we initially separated 10% of the dataset to form an independent test set for final evaluation. The remainder was then divided, reserving 10% for validation purposes. This partitioning strategy resulted in 3,523 training samples, 392 validation samples, and 436 test samples, ensuring that our model was both trained and validated on diverse subsets of the data, promoting generalizability and robust performance across unseen datasets. The age distribution for train, validation and test sets were shown in Fig. 1c, 1d, 1e respectively.

Model Architecture

Temporal Convolutional Networks (TCNs) excel in sequence modeling tasks due to their hierarchical architecture, which adeptly manages long-range dependencies. Our implementation incorporates key architectural features that improve the efficiency and accuracy of the network in age prediction from the DNA methylation data shown in Fig. 2.

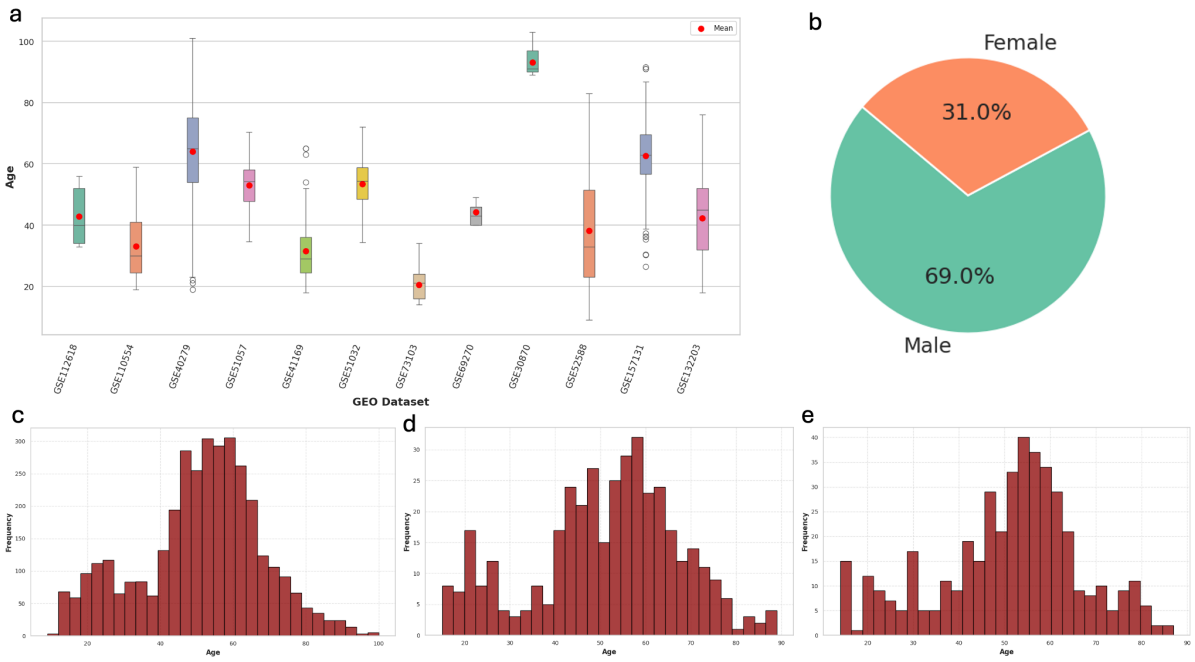

Figure 1: Analysis of GEO Datasets for Age Prediction. a) Age Distribution: A bar plot illustrating the age range from 0 to over 100 years across 12 GEO datasets, with annotations for mean and standard deviation to highlight the diversity in age representation. b) Gender Proportion: Pie chart showing the gender ratio of the sampled populations, restricted to datasets with available sex information. c-e) Data Partitioning: Histograms depicting the age distribution within the training, validation, and test sets, demonstrating consistent distribution patterns across different dataset partitions.

Temporal Block

The fundamental unit of our TCN, the TemporalBlock, consists of a series of convolutional layers coupled with ReLU activation (Agarap 2018) and dropout for regularization. Each block features a residual connection, streamlining the training of deep networks by preserving gradient flow and mitigating the vanishing gradient problem. This design ensures robust feature extraction across different layers, facilitating the capture of complex patterns in methylation profiles. Each *TemporalBlock* in our TCN processes the input through two main convolutional layers, each followed by a ReLU activation and dropout, and includes a residual connection:

$$x_{\text{out}} = \text{Dropout}(\text{ReLU}(\text{Conv}(x_{\text{in}}))) + x_{\text{res}} \tag{1}$$

where x_{res} is the residual connection that may involve a transformation if the dimensions do not match:

$$x_{\text{res}} = \begin{cases} \text{Conv}_{1\times1}(x_{\text{in}}) & \text{if shape mismatch,} \\ x_{\text{in}} & \text{otherwise.} \end{cases} \tag{2}$$

Dilated Convolutions

To expand the model's receptive field without a proportional increase in parameters, we employ dilated convolutions. This approach allows the network to integrate information over larger expanses of the input sequence, capturing the distant relationships between CpG sites that are crucial for accurate age estimation. The dilation factor increases exponentially with each subsequent layer, enhancing the model's ability to assimilate broader contextual information from the methylation data:

$$y(t) = \sum_{i=0}^{N-1} f(i) \cdot x(t - s \cdot i) \tag{3}$$

where s is the dilation factor, N is the filter size, f is the filter, and x is the input.

Final TCN Prediction

The overarching model structure, TCNModel, stacks multiple TemporalBlock layers, each refining the feature representations extracted from the data. The architecture concludes with a linear layer that transforms the high-level features into a final age prediction. This layer acts as a regression output, providing a quantitative estimate of biological age:

$$\text{Age} = \text{Linear}(\text{GlobalAvgPool}(x_{\text{final}})) \tag{4}$$

We integrate batch normalization to stabilize the learning process, leading to faster convergence and enhanced training dynamics. Dropout is strategically placed within the TemporalBlocks to prevent overfitting, ensuring that our model generalizes well to new, unseen data. The configuration of these components within the TCN framework allows for a powerful, yet efficient, approach to modeling the intricate

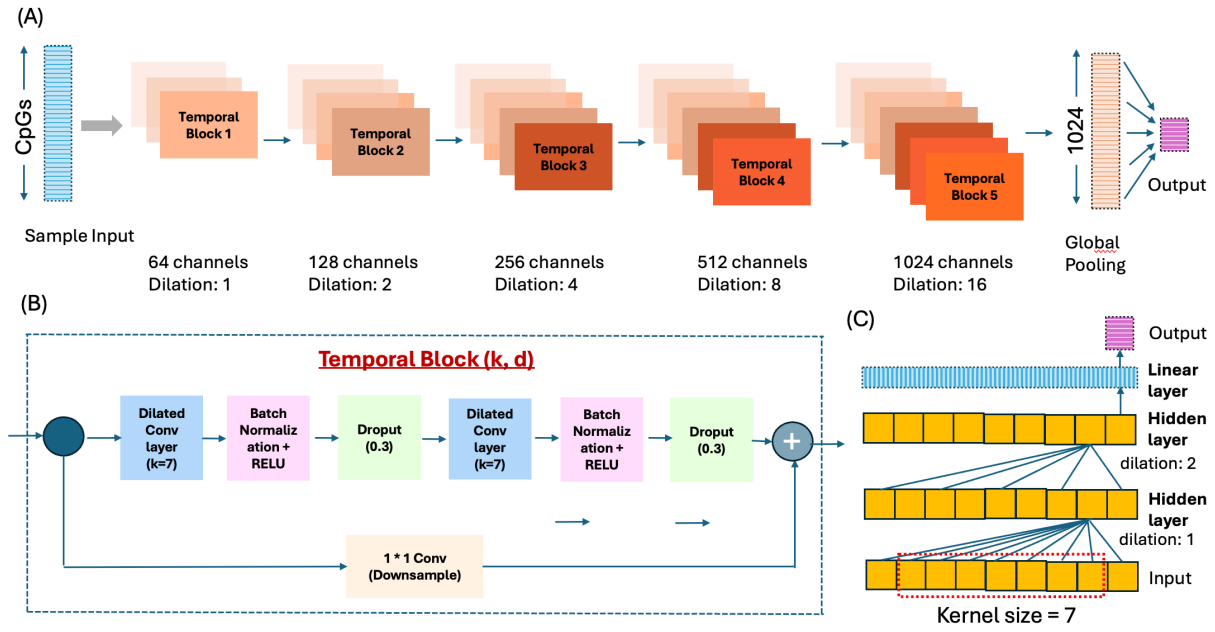

Figure 2: DeepAge Model Architecture. (A)Input and TCN Overview: Showcases an input sample with CpG features processed through a five-layer TCN. Each layer doubles the channel count from 64, with exponentially increasing dilations and adaptive padding, culminating in a global max pooling layer that reduces all features to a 1024-dimensional embedding for the output prediction. (B) Temporal Block Structure: Details the configuration of a temporal block, including dual convolutional layers followed by batch normalization, ReLU, and dropout, complemented by a skip connection for enhanced gradient flow. (C) Dilation mechanics: Highlights the role of increasing dilation in the capture of distant CpG interactions.

relationships inherent in epigenetic data, ultimately leading to more precise age predictions based on DNA methylation patterns.

Experimental Setup and Model Training

In this study, we developed a temporal convolutional network (TCN) to predict epigenetic age. Our model architecture included five residual blocks, each comprising two convolutional layers followed by batch normalization and ReLU activation functions. To mitigate overfitting, we introduced a dropout rate of 30% after each convolutional layer. The network architecture is designed such that each layer maintains the sequence length of the input (e.g., 57 CpGs), using padding calculated to accommodate the dilation effects. This design allows the number of channels to increase with each layer, enhancing the network's ability to learn increasingly complex features at deeper levels.

The final layer of the model processes the output with an embedding size that expands to 1024 channels, while preserving the original sequence length. This setup enables the model to represent more complex features without altering the temporal resolution of the input. Following the last temporal block, we applied global average pooling across the sequence dimension, condensing the temporal information into a single vector per feature channel. Consequently, each of the 1024 channels represents the average feature value across all time steps. A linear layer subsequently reduces this 1024-dimensional vector to a single predictive output, focusing on the most critical features for age prediction. This method ensures that the model prioritizes the most significant features extracted throughout the sequence, thereby enhancing its robustness and adaptability to various input lengths.

For training, we used a batch size of 32, used the Adam optimizer with a learning rate of 0.001 and adopted the mean square error as the loss function. The training process was designed to run for up to 200 epochs, with a patience parameter of 5 set for early stopping based on validation loss. The training loss curve is shown in the Fig. 3b together with the validation loss, and the training performance is shown in Table 1 along with the validation and test results. This strategy was implemented to prevent overfitting and stop training when the model ceased to show improvements in the validation data set.

Metric	Train	Val	Test
Mean Abs Error (MAE)	4.18	4.92	4.88
Coeff. of Determination (R^2)	0.89	0.85	0.84
Mean Abs Deviation (MAD)	13.19	12.98	12.23
Root Mean Sq Error (RMSE)	5.34	6.4	6.21
Median Abs Error (MedAE)	3.52	3.86	3.98

Table 1: Performance of DeepAge on training, validation, and test datasets (using 184 CpGs)

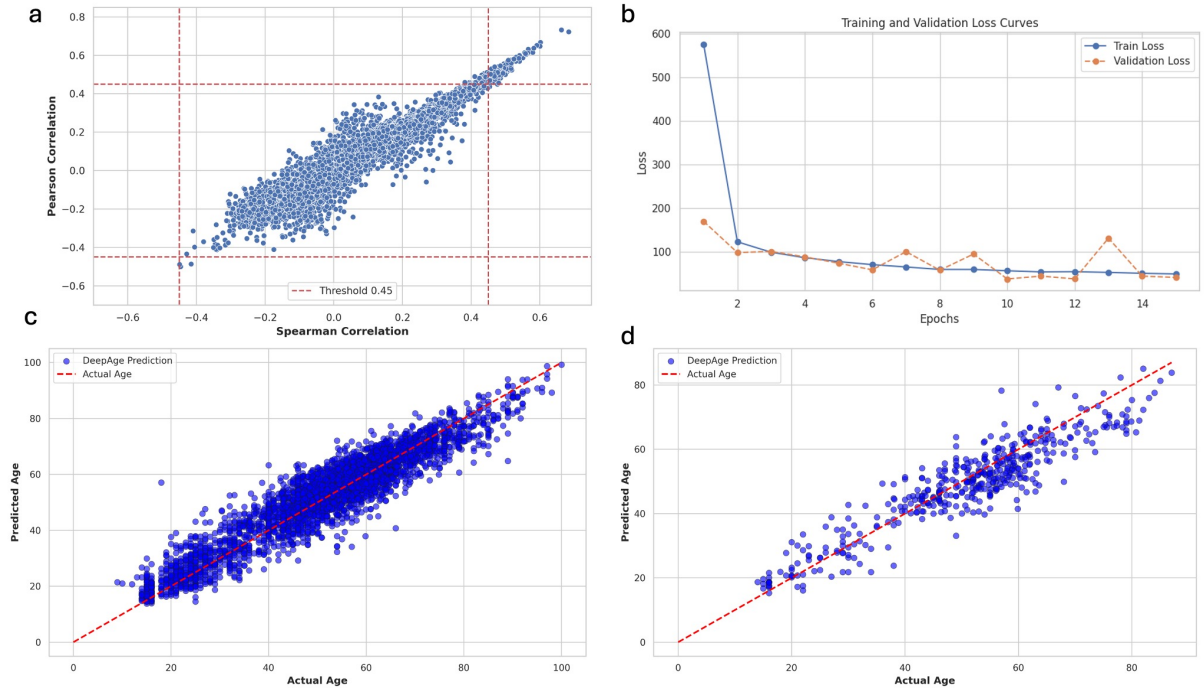

Figure 3: Feature Selection and Model Training. a) Dual Correlation Method: Depicts feature selection using a 0.45 correlation threshold. b) Loss curves: Shows training and validation loss curves. c-d) Age predictions: Displays predicted vs. actual age plots for both training and test sets.

Evaluation

To effectively measure different model's age prediction capabilities from DNA methylation data, we employed a robust and comprehensive set of regression metrics. These include:

- Mean Absolute Error (MAE)

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{5}$$

- Median Absolute Error (MedAE)

$$\text{MedAE} = \text{median}(|y_1 - \hat{y}_1|, \ldots, |y_n - \hat{y}_n|) \tag{6}$$

- R-Squared (R^2)

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \tag{7}$$

- Mean Absolute Deviation (MAD)

$$\text{MAD} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \text{median}(Y)| \tag{8}$$

These metrics facilitate a nuanced assessment of prediction accuracy. MAE and RMSE directly reflect average errors in age estimates, with RMSE adding weight to larger discrepancies. R-squared offers insight into how much age-related variance our model captures compared to the baseline model. Collectively, these metrics underscore our model's ability to generalize well across diverse methylation profiles, substantiating its efficacy in biological age estimation.

Model	#CpGs	MAE	R^2	RMSE	MedAE
Hannum	6	34.16	-4.63	37.31	32.07
Horvath1	297	40.56	-6.61	43.34	38.55
Horvath2	100	39.58	-6.34	42.58	37.35
PhenoAge	468	40.04	-5.77	40.87	39.57
Lin	89	45.69	-7.77	46.55	46.63
DunedinPACE	4	40.38	-5.76	43.33	38.13
YingAdaptAge	53	58.83	-13.93	60.74	56.71
YingDamAge	50	28.31	-3.19	32.18	25.58
XGBoost	57	23.77	-2.22	28.21	20.42
Random Forest	57	11.57	0.12	14.77	9.03
CNN-Attention	57	11.07	0.18	14.24	8.48
CNN (3 layers)	57	9.26	0.49	11.25	8.43
LSTM (2 layers)	57	7.14	0.68	8.95	6.05
DeepAge	57	**5.55**	**0.79**	**7.16**	**4.27**

Table 2: Performance comparison of different state-of-the-art models on the test dataset

Results

Comparison of Different Machine Learning and Deep Learning Based Methods on Age Prediction

In our study, we evaluated the performance of various age prediction methodologies, comparing traditional machine learning techniques, ensemble methods, and advanced deep learning architectures against our DeepAge model shown in Table 2. These included regression approaches, gradient boosting, stacking, deep neural networks such as CNN (O'shea and Nash 2015), LSTM (Graves and Graves 2012),

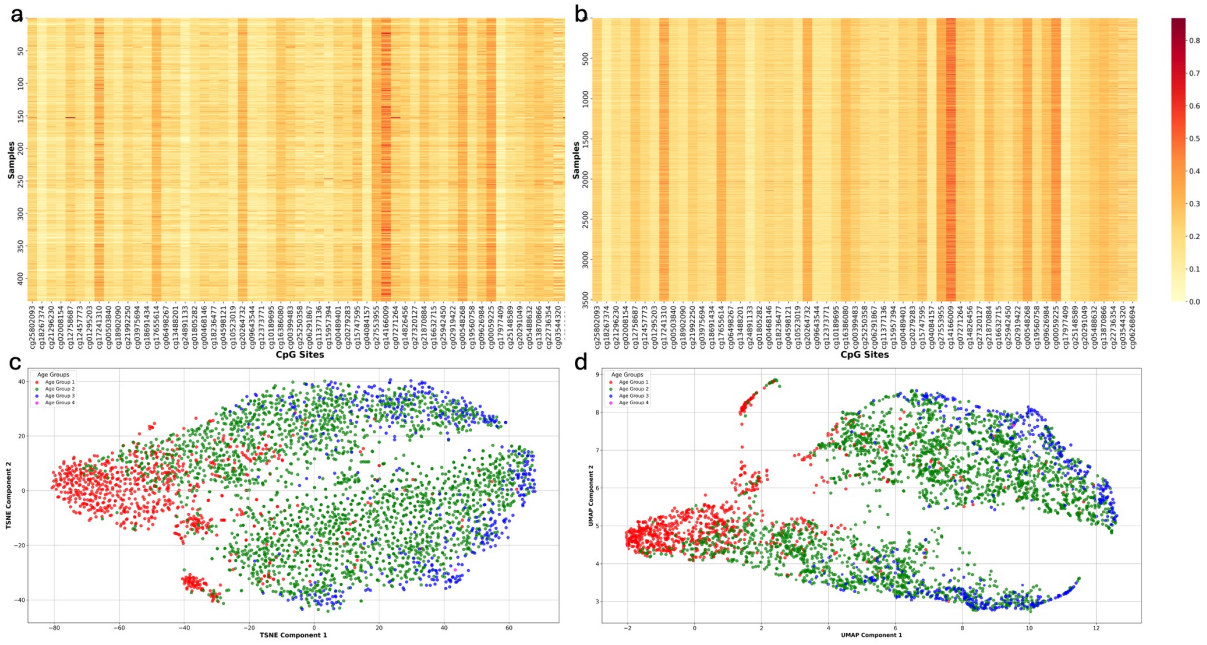

Figure 4: Visualization of the methylation data. a-b) Heatmaps of CpG sites display methylation levels across samples for test and training sets, respectively. c-d) Dimensionality Reduction: Illustrates t-SNE and UMAP visualizations of methylation data, categorized into four age groups (0-25, 25-50, 50-75, 75-100) showing distinct patterns for age groups.

and CNN combined with attention mechanisms (Vaswani et al. 2017). For regression based methods and some other models, we utilized the CpG coefficients from the Biolearn library for this comparative analysis, selecting common CpGs between the provided lists and our dataset to facilitate age prediction. We developed various machine learning and deep learning models from scratch, including DeepAge, which outperformed others in using 57 CpGs, as shown in the comparison of predicted results versus actual ages using 184 CpGs in the Fig. 3c, 3d for the train and test sets, respectively. This success highlights the benefit of treating methylation data as sequential patterns, enhancing predictive accuracy. Specifically, DeepAge's use of Temporal Convolutional Networks (TCN) with residual blocks and dilated convolutions significantly improves age prediction. Residual blocks preserve gradient strength during backpropagation for deeper networks without performance loss, while dilated convolutions broaden the receptive field, capturing distant CpG interactions critical for detecting complex, age-predictive methylation patterns.

Conversely, convolutional approaches were moderately effective, indicating that capturing long-range CpG interactions could improve accuracy. Traditional methods like random forest and XGBoost underperformed, likely unable to capture complex CpG interactions. Other epigenetic clocks, based on Biolearn coefficients, also performed poorly, possibly due to less relevant CpG selections compared to those in our more precisely curated model. This analysis underscores DeepAge's robustness and effectiveness in using epigenetic sequencing for age prediction, establishing a new accuracy benchmark in the field.

Metric	Threshold & #CpGs				
	T=0.40 (407)	**T=0.45 (184)**	T=0.50 (57)	T=0.55 (16)	T=0.60 (5)
MAE	5.2	**4.88**	5.55	6.72	7.69
R^2	0.82	**0.84**	0.79	0.69	0.60
RMSE	6.62	**6.21**	7.16	8.62	9.88
MedAE	4.32	**3.98**	4.27	5.24	5.99

Table 3: Effect of Number of CpGs on Age Prediction

Effect of Number of CpG Sites and Their Correlation on Age Prediction

The performance of age prediction models is significantly affected by the number of CpG sites used. More CpGs increase genomic coverage and capture a broader range of aging-related methylation patterns, but also add complexity and risk of overfitting, making the model perform well on training data but poorly on new data. Additionally, a larger set of features makes it harder to identify the most influential CpGs, complicating the interpretation of results and biological validations. Effective feature selection is essential to identify the most informative CpGs, improving model performance and computational efficiency.

To determine the optimal number of CpG sites among our samples, we evaluated five different CpG sets selected using Dual-Correlation feature selection at varying thresholds (0.45, 0.50, 0.55, 0.60) shown in Table 3. We found that models with a moderate number of highly associated CpGs perform well, but performance declines beyond a certain point due to overfitting. For example, using 184

Model Parameters	MAE	R^2	RMSE	MedAE
TCN k=3, 3 layers, drop=0.2, n_channels = [64, 128, 256]	9.24	0.47	11.41	7.93
TCN k=7, 3 layers, BN, drop=0.3, n_channels = [64, 128, 256]	6.47	0.73	8.1	5.45
TCN k=7, 5 layers, BN, drop=0.3, n_channels = [64, 128, 256, 512, 1024]	**5.55**	**0.79**	**7.16**	**4.27**
TCN k=7, 5 layers, BN, drop=0.3, n_channels = [64, 128, 256, 512, 512, 1024, 1024, 2048]	12.8	-0.03	15.93	10.17

Table 4: Ablation study on model parameters of DeepAge (57 CpGs)

CpGs yielded better results than using 407 CpGs, indicating the onset of overfitting. In our analysis, a set of 57 CpGs emerged as a balanced choice, providing robust performance while minimizing complexity.

Visualization of CpG Site Methylation Levels Patterns Across Samples

In our study, we meticulously analyzed the methylation patterns across various CpG sites to gain insight into their implications for biological age estimation. The heatmap provided visually demonstrates the variation in methylation levels across different samples. Each row represents a sample, and each column corresponds to a specific CpG site. The color gradient, which varies from light yellow (low methylation) to deep red (high methylation), facilitates the identification of CpG sites with pronounced methylation changes. From this visualization, we can discern clear patterns of methylation across specific sites, highlighting regions with potential biological significance in aging processes. Sites with consistently higher or lower methylation across samples could indicate key regulatory regions impacting gene expression tied to aging. This methodical mapping enables us to target these significant CpG sites for deeper analysis, potentially guiding further experimental investigations. In addition, the heat map helps to identify outliers and trends that may not be evident through numerical data alone. In the Fig. 4a, 4b we showed the methylation levels for each sample in the test and training set, respectively, across all samples. For example, cg14166009, cg00059225 is highly methylated for both the train and test samples. By correlating these patterns with age groups and other phenotypic data, we can better understand the role of epigenetic modification in aging and develop more accurate predictive models for biological age. This approach not only enhances the precision of age estimation models but also enriches our understanding of the epigenetic mechanisms that underlie age-related changes.

Ablation Study on Model Parameters of DeepAge (57 CpGs)

To refine our age prediction model, DeepAge, we performed ablation experiments to determine the optimal architectural features, including layer count, kernel size, and regularization methods like batch normalization and dropout. The findings, summarized in Table 4, illustrate the effectiveness of our design, particularly with the 57 CpG model configuration which demonstrates DeepAge's ability to deliver accurate age predictions from DNA methylation profiles. Our results indicated that a five-layer model with a kernel size of 7 best captures relevant CpG site information, enhancing pattern detection for aging. Incorporating batch normalization and a dropout rate of 0.3 improved the model's generalizability and robustness, effectively reducing overfitting and ensuring consistent performance across datasets.

Analysis of Variation in DNA Methylation and Age Patterns Across Different Age Groups Through Dimensionality Reduction Techniques

In our study, we used advanced dimensionality reduction techniques, specifically t-SNE (Van der Maaten and Hinton 2008) and UMAP (McInnes, Healy, and Melville 2018), to analyze DNA methylation data in relation to epigenetic age. Our results, displayed in Fig. 4c, 4d categorize the samples into four age groups (0-100 years) and reveal notable differences in the methylation patterns. The visualization of t-SNE (Fig. 4c) shows some age-based clustering, though with significant overlap, indicating a less defined age-related structure. In contrast, UMAP (Fig. 4d) demonstrates clearer delineation of age groups, suggesting its effectiveness in capturing the global structure and biological significance of age-related changes in methylation. These findings underscore the potential of methylation patterns as biomarkers of biological aging and highlight the utility of UMAP in epigenetic research to understand the effects of age on methylation.

Discussions

In this study, we developed DeepAge, a deep learning framework that estimates epigenetic age using DNA methylation data. Utilizing Temporal Convolutional Networks (TCNs), DeepAge captures long-range dependencies between CpG sites, outperforming traditional epigenetic clocks. The model's architecture features dilated convolutions to expand the receptive field efficiently, supported by residual connections and dropout, ensuring robust learning with limited data. However, our study has limitations that suggest areas for future research. It is limited to 12 human blood sample datasets from the BioLearn library, which restricts the diversity and scope of our epigenetic aging analysis. Future research could broaden this by including more varied datasets and expanding the model to multiple tissue types, enhancing prediction accuracy and applicability. Despite these limitations, DeepAge represents a significant advancement in epigenetic age estimation, highlighting deep learning's potential in biomedical research and setting the stage for further improvements.

Code and Data Availability

Code is available at https://github.com/Sajib-006/DeepAge We have used publicly available data from the NCBI GEO database (https://www.ncbi.nlm.nih.gov/gds) from biolearn library (https://bio-learn.github.io/data.html)

Acknowledgments

The authors thank Virginia Tech for computational resources, the Department of Computer Science for ongoing support and facilities, and GEO and Biolearn database .

References

Agarap, A. F. 2018. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*.

Belsky, D. W.; Caspi, A.; Corcoran, D. L.; Sugden, K.; Poulton, R.; Arseneault, L.; Baccarelli, A.; Chamarti, K.; Gao, X.; Hannon, E.; et al. 2022. DunedinPACE, a DNA methylation biomarker of the pace of aging. *Elife*, 11: e73420.

Breiman, L. 2001. Random forests. *Machine learning*, 45: 5–32.

Chen, T.; and Guestrin, C. 2016. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794.

De Winter, J. C.; Gosling, S. D.; and Potter, J. 2016. Comparing the Pearson and Spearman correlation coefficients across distributions and sample sizes: A tutorial using simulations and empirical data. *Psychological methods*, 21(3): 273.

Edgar, R.; Domrachev, M.; and Lash, A. E. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic acids research*, 30(1): 207–210.

Graves, A.; and Graves, A. 2012. Long short-term memory. *Supervised sequence labelling with recurrent neural networks*, 37–45.

Hannum, G.; Guinney, J.; Zhao, L.; Zhang, L.; Hughes, G.; Sadda, S.; Klotzle, B.; Bibikova, M.; Fan, J.-B.; Gao, Y.; et al. 2013. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular cell*, 49(2): 359–367.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Horvath, S. 2013. DNA methylation age of human tissues and cell types. *Genome biology*, 14: 1–20.

Horvath, S.; Oshima, J.; Martin, G. M.; Lu, A. T.; Quach, A.; Cohen, H.; Felton, S.; Matsuyama, M.; Lowe, D.; Kabacik, S.; et al. 2018. Epigenetic clock for skin and blood cells applied to Hutchinson Gilford Progeria Syndrome and ex vivo studies. *Aging (Albany NY)*, 10(7): 1758.

Lea, C.; Vidal, R.; Reiter, A.; and Hager, G. D. 2016. Temporal convolutional networks: A unified approach to action segmentation. In *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14*, 47–54. Springer.

Levine, M. E.; Lu, A. T.; Quach, A.; Chen, B. H.; Assimes, T. L.; Bandinelli, S.; Hou, L.; Baccarelli, A. A.; Stewart, J. D.; Li, Y.; et al. 2018. An epigenetic biomarker of aging for lifespan and healthspan. *Aging (albany NY)*, 10(4): 573.

Li, L.; Zhang, C.; Liu, S.; Guan, H.; and Zhang, Y. 2021. Age prediction by DNA methylation in neural networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(3): 1393–1402.

Lin, Q.; Weidner, C. I.; Costa, I. G.; Marioni, R. E.; Ferreira, M. R.; Deary, I. J.; and Wagner, W. 2016. DNA methylation levels at individual age-associated CpG sites can be indicative for life expectancy. *Aging (Albany NY)*, 8(2): 394.

McInnes, L.; Healy, J.; and Melville, J. 2018. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*.

O'shea, K.; and Nash, R. 2015. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*.

Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. 2011. Scikit-learn: Machine learning in Python. *Journal of machine learning research*, 12(Oct): 2825–2830.

Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; and Salakhutdinov, R. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1): 1929–1958.

Van der Maaten, L.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9(11).

Varshavsky, M.; Harari, G.; Glaser, B.; Dor, Y.; Shemer, R.; and Kaplan, T. 2023. Accurate age prediction from blood using a small set of DNA methylation sites and a cohort-based machine learning algorithm. *Cell Reports Methods*, 3(9).

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Wang, H.; Cai, R.; Zong, X.; He, Z.; and Zhang, L. 2023. MSCAP: DNA Methylation Age Predictor based on Multiscale Convolutional Neural Network. In *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 3582–3586. IEEE.

Ying, K.; Paulson, S.; Perez-Guevara, M.; Emamifar, M.; Martínez, M. C.; Kwon, D.; Poganik, J. R.; Moqri, M.; and Gladyshev, V. N. 2023. Biolearn, an open-source library for biomarkers of aging. *bioRxiv*, 2023–12.

Yu, F.; and Koltun, V. 2015. Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*.

Zhao, J.; Li, H.; Qu, J.; Zong, X.; Liu, Y.; Kuang, Z.; and Wang, H. 2024. A multi-organization epigenetic age prediction based on a channel attention perceptron networks. *Frontiers in Genetics*, 15: 1393856.